

Claims

1. Method for training at least one learning-capable system comprising the steps of:

providing a predetermined training data set corresponding to a predetermined number of subjects comprising a predetermined input data set and a predetermined outcome data set,

augmenting the input data set and/or the outcome data set, and

training each learning-capable system using the augmented input data set and/or the augmented outcome data set.
2. Method according to claim 1, wherein the augmenting step comprises the steps:

estimating propensity score data for each subject depending on its input data,

dividing the propensity score data into at least two strata,

assigning each subject to a stratum, and

augmenting the input data of each subject by its propensity score data and/or its stratum data.
3. Method according to claim 2, wherein the training step comprises the step of optimizing the operating point parameters for each stratum.
4. Method according to claim 3, wherein the operating point parameters are optimized such that the median of all output data vanishes for each stratum.
5. Method according to one of the preceding claims, wherein the augmenting step comprises the step of:

generating a plurality of augmented training data sets by augmenting the input data set using a predetermined statistical model.
6. Method according to claim 5 for training at least two learning-capable systems, wherein the training step comprises the steps of:

training each learning-capable system using a subset of the plurality of augmented training data sets,

constructing scores for each outcome for each trained learning-capable system, and
determining characteristics of the distributions of the scores for each subject.

7. Method according to claim 5 or 6, wherein the input data set is augmented using a generalized Markov chain Monte-Carlo method.

8. Method according to one of the preceding claims, wherein the augmenting step comprises the steps of:

providing a further learning capable-system and a further predetermined training data set comprising a further predetermined input data set and a further predetermined outcome data set,

training the further learning-capable system using the further predetermined training data set, and

augmenting the input data set by at least one additional input variable taken from the further input data, further outcome data and/or internal output data obtained from the trained further learning-capable system.

9. Method according to claim 8, wherein the additional input variables comprise all further input data and all further outcome data of a subset of subjects of the further training data set.
10. Method according to one of the preceding claims, wherein the outcome data of the training data set is time-dependent and the augmenting step comprises pre-transforming the time variable of the training data set in such a way that the hazard rate with respect to a predetermined outcome is a predetermined function of the time variable.
11. Method for using a learning-capable system trained according to the method of one of the preceding claims using the input data of a subject, characterized in that the outcome is corrected with respect to a predetermined reference subject.

12. Method for using at least two learning-capable systems trained according to the method of one of the claims 7 – 10 in combination with claim 6 using the input data of a subject, comprising the steps of:

presenting the input data of the subject to each of the learning-capable systems and

constructing a score for the output data obtained from the learning-capable systems.
13. Method to create a composite training data set, in particular for training a learning-capable system according to the method one of the claims 1 – 10, comprising the steps of:

providing an aggregated evidence data set,

disaggregating the aggregated evidence data set to obtain a disaggregated training data set with virtual subjects, and

merging the disaggregated training data set with a further training data set.
14. Method according to claim 13, wherein the merging step comprises the step of choosing a real training data set as the further training data set.
15. Method according to claim 13 or 14, wherein the disaggregation step comprises the step of assigning at least a value of one auxiliary variable to each virtual subject of the disaggregated training data set.
16. Method according to one of the claims 1 – 10, wherein the predetermined training data set is provided using the method according to one of the claims 13 – 15.
17. A computer program product directly loadable into the internal memory of a digital computer, comprising software code portions for performing the steps of the method of one of the preceding claims, when said product is run on a computer.
18. A computer program product stored on a medium readable by a computer, comprising computer readable program means for causing a computer to perform the steps of the method of one of the preceding claims, when said product is run on a computer.